AD-A041 985   NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS   F/G 12/1
             A STUDY OF POWER TRANSFORMS.(U)
             MAY 77   R A FONDREN   AF-AFOSR-2796-75
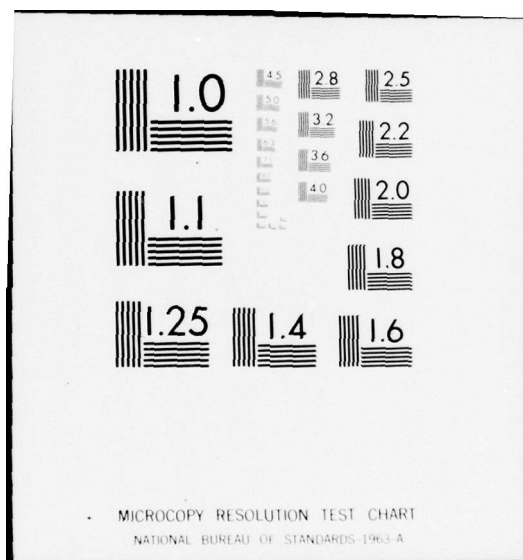
UNCLASSIFIED   MIMEO SER-1123   AFOSR-TR-77-0712   NL

| OF |
ADA041985

END

DATE
FILMED

8 - 77

1.0

2.8   2.5

3.2   2.2

3.6

4.0   2.0

1.1

1.8

1.25   1.4   1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY OF NORTH CAROLINA
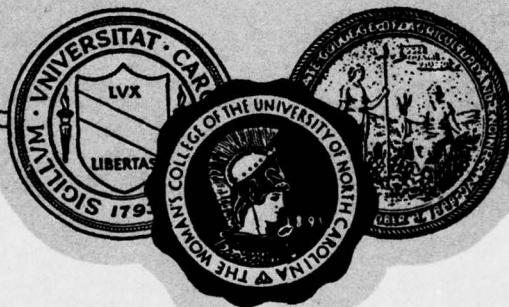
A STUDY OF POWER TRANSFORMS

Robert A. Fondren

DDC

JUL 26 1977

C

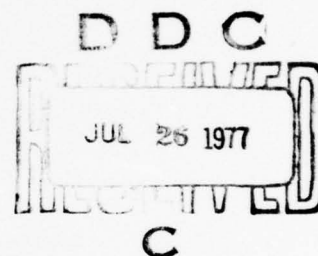## DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

A STUDY OF POWER TRANSFORMS*

by

Robert A. Fondren

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1123

May, 1977

(See 1473)

CHAPTER I

INTRODUCTION

In most practical situations, the data one has in hand
do not follow the normal distribution; if the data were nor-
mal, one could use all of the procedures in multivariate
analysis which apply to normally distributed data. In this
paper, a routine method for making the data follow the nor-
mal distribution more closely is discussed; this method is
to transform the data.

The idea of using certain transformations before
analyzing data is well known. Two of the most common trans-
formations which are usually introduced in an elementary
analysis of variance course are the log transformation and
the reciprocal transformation. These are special cases of
the family of power transformations which was introduced in
a paper by Box and Cox (1964). The family of transformations
is given by

$$y^{(\lambda)} = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \\ \log y & \lambda = 0 \ . \end{cases}$$

Several methods are available for estimating the value of λ
which will make the data most closely normal. These esti-
mators and related tests of the hypothesis $\lambda = \lambda^0$ are dis-

cussed in the second part of the paper.

Also included is a summary of the analysis of two experiments which are introduced in the paper by Box and Cox. The power of the tests is studied briefly, using the results given in an example found in a paper by Atkinson (1973).

In the third section we discuss the influence curve of an estimator, which is an indicator of the influence of an observation on the estimator. The influence curves of several robust estimators of location are given, to introduce the concept. Then the idea of the influence curve is applied to the estimator of λ .

Another purpose of the paper is to further study the power of the tests which are introduced in the second section. This study of the power follows the method which was outlined in the example in the paper by Atkinson. This example, which uses the biological data from the Box and Cox paper, is used as the basis for a series of simulations. These simulations are performed using data other than normally distributed, to indicate the robustness of the tests, as well as the power of the tests under distributions other than the normal. The results of these simulations are given later in the paper.

CHAPTER II

ESTIMATES OF THE BEST TRANSFORMATION

Box and Cox (1964) worked with the observations $y_1, \ldots, y_n$, arranged in a nx1 vector $\underline{y}$, and defined the family of transformations

$$\underline{y}^{(\lambda)} = ((y_i^{(\lambda)})) = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log y_i & \text{if } \lambda = 0 \end{cases} \tag{1}$$

where λ is a parameter to be estimated. They assumed that for some λ the transformed observations $\underline{y}^{(\lambda)}$ would be normally distributed and would satisfy the linear model

$$E[\underline{y}^{(\lambda)}] = X \underline{\beta}$$

with constant error variances and absence of interactions. Here X is a matrix of constants and $\underline{\beta}$ is a vector of parameters.

The problem that we wish to examine is the estimation of λ. In the paper by Box and Cox, the method of maximum likelihood was used to find the value of λ for which the data would most closely follow a normal model with constant variances and no interactions.

If $\underline{y}^{(\lambda)}$ is the vector of transformed observations,

the maximized log likelihood is

$$Lmax(\lambda) = -\tfrac{1}{2} n \log \hat{\sigma}^2(\lambda) + (\lambda-1) \Sigma \log Y_i \qquad (2)$$

where $\hat{\sigma}^2(\lambda) = \dfrac{1}{n} \underline{y}(\lambda)' A\underline{y}(\lambda)$ is an estimate of the error variance σ^2 and $A = I - X(X'X)^{-1}X'$.

Using this maximized log likelihood, there are two ways by which an estimate of λ can be found. One way is to plot Lmax (λ) versus λ , find the point which maximizes the likelihood and choose that point to be the estimate $\hat{\lambda}$ of λ .

Another way to find $\hat{\lambda}$ is to take the derivative of the likelihood function, equate it to zero and thus find the estimate of λ. This method should be used if more precision is desired than that given by plotting the likelihood function. The derivative of Lmax (λ) with respect to λ is

$$\frac{d}{d\lambda} Lmax(\lambda) = n \frac{\underline{y}(\lambda)' A\underline{u}(\lambda)}{\underline{y}(\lambda)' A\underline{y}(\lambda)} + \frac{n}{\lambda} + \Sigma \log y_i \qquad (3)$$

or multiplying through by $\underline{y}(\lambda)' A\underline{y}(\lambda)$,

$$= -n \underline{y}(\lambda)' A\underline{u}(\lambda) + \underline{n} \underline{y}(\lambda)' A\underline{y}(\lambda) + (\underline{y}(\lambda)' A\underline{y}(\lambda)) \Sigma \log y_i$$

to give a form which is easier to use. On equating this to zero, we would obtain the maximum likelihood estimate of λ, where $\underline{u}(\lambda)$ is the vector with components $(\frac{1}{\lambda} y_i^{\lambda} \log y_i)$.

To simplify the problem we could consider the normalized observations

$$\underline{z}(\lambda) = \underline{y}(\lambda) / J^{1/n} \qquad (4)$$

where $J = J(\lambda ; \underline{y}) = \prod_{i=i}^{n} \left| \frac{dy_i(\lambda)}{d\lambda} \right|$.

Using the transformations (1), we now get

$$Lmax(\lambda) = \tfrac{1}{2} n \log \hat{\sigma}^2(\lambda ; \underline{z})$$

where $\hat{\sigma}^2(\lambda;\underline{z}) = \underline{z}^{(\lambda)'}A\underline{z}^{(\lambda)}/n$.

Here $\underline{z}^{(\lambda)} = ((z_i^{(\lambda)})) = \dfrac{y_i^{\lambda}-1}{\lambda\dot{y}^{\lambda-1}}$ (5)

where $\dot{y}$ is the geometric mean of the observations.

Box and Cox considered two examples in their paper, a biological example and a textile example. The biological example was a 3 x 4 factorial experiment with the two factors poisons and treatments. They used the transformations (1) and found the value $\hat{\lambda} = -0.75$ to be the maximum likelihood estimate of λ. Thus the familiar reciprocal transformation with $\lambda = -1$ was used, because the results were easier to analyze using this transformation and because customarily a more familiar transformation should be used. The textile example was a 3^3 experiment. For this example they considered the normalized observations $\underline{z}^{(\lambda)}$ and found the maximum likelihood estimate of λ to be $\hat{\lambda} = -0.06$. Thus for this example they used the log transformation with $\lambda = 0$, because the data was easier to analyze and because the log transformation is more familiar.

One problem with the method used by Box and Cox is that they assumed that for some λ, the transformed observations followed a normal model with constant error variances and no interactions. It is hard to believe that for any data vector $\underline{y}$ to be considered, there exists a value of λ which satisfies these conditions exactly.

A paper by Draper and Cox (1969) addressed this problem. In their paper, the conjecture was that even if these

three conditions could not be satisfied as exactly for any λ, the estimate of λ found by the Box and Cox method might still be of interest. They showed by example that for this estimate of λ, the resulting distribution was close enough to normal to be useful. One problem they found was that the sample size would have to be quite large before the resulting estimate of λ was precise. This precision was measured by finding the variance of the estimate $\hat{\lambda}$.

Another paper in which the Box and Cox transformations were considered was a paper by Schlesselman (1971). In his paper, he stated that the maximum likelihood estimate was not invariant under scaling of the original observations from $\underline{y}$ to $\omega\underline{y}$ unless the X matrix from the linear model contained a column of ones or, in other words, an additive constant could be removed from the model. In most practical situations, the model is defined in such a way that the X matrix does allow for removal of an additive constant; therefore, this problem is not important.

In a paper by Atkinson (1973), the Box and Cox maximum likelihood test statistic was expressed in the form

$$T_L = [\ -2\ \{Lmax(\lambda_0) - Lmax(\hat{\lambda})\}\ ]^{\frac{1}{2}} \tag{6}$$

where $\hat{\lambda}$ is the estimate of λ from the maximum likelihood method and λ_0 is the exact value of λ. Since T_L^2 has an asymptotic χ_1^2 distribution, the statistic T_L has a standard normal asymptotic distribution. Since this statistic has an asymptotic normal distribution, a test was desired which

would have an exact distribution. Exact tests could thus be
made instead of tests depending on an asymptotic distri-
bution.

Andrews, in a 1971 paper, derived an exact test
statistic for testing $\lambda = \lambda_0$ which also had the advantage of
being computationally simpler than the maximum likelihood
statistic. To derive his test statistic, Andrews started
with the transformations (1) and assumed that for some λ ,
the vector of transformed observations $\underline{y}^{(\lambda)}$ could be ex-
pressed in the form

$$\underline{y}^{(\lambda)} = X\underline{\beta} + \underline{e} \tag{7}$$

where X and $\underline{\beta}$ are as defined previously and $\underline{e}$ is the vector
of errors, with mean 0 and variance σ_y^2.

He assumed that the values $\underline{y}^{(\lambda_0)}$, which are the trans-
formed observations at the true value of λ , follow a Taylor
expansion about λ_0 given by

$$\underline{y}^{(\lambda_0)} = X\underline{\beta} + \underline{v}(\lambda - \lambda_0) + \underline{e} \tag{8}$$

where the remainder terms in higher powers of λ were ig-
nored. The vector $\underline{v}$ was defined by

$$\underline{v} = ((v_i)) = \left[\frac{\partial y_i(\lambda)}{\partial \lambda} \right]_{\lambda = \lambda_0}. \tag{9}$$

The vector $\underline{v}$ depends on $\underline{y}$, so we must somehow modify this
vector to construct a test statistic. This was accomplished
by calculating the vector $\hat{\underline{v}}$ defined by

$$\hat{\underline{v}} = ((\hat{v_i})) = \left[\frac{\partial y_i(\lambda)}{\partial \lambda} \right]_{\lambda = \lambda_0, \underline{y} = \hat{\underline{y}}} \tag{10}$$

where the values $\hat{\underline{y}}$ are the fitted values of $\underline{y}$, given by $\hat{\underline{y}} = (X'X)^{-1}X'\underline{y}^{(\lambda)}$. The test statistic was then derived using the method given in Milliken and Graybill (1970). The resulting statistic was

$$T_A = - \frac{\underline{y}^{(\lambda)'} A\hat{\underline{v}}}{s_y \, (\hat{\underline{v}}'A\hat{\underline{v}})^{\frac{1}{2}}} \tag{11}$$

(as expressed in Atkinson) where s_y^2 is an estimate of the error variance σ_y^2. This statistic has an exact t distribution; it would have a standard normal distribution if the variance σ^2 were known.

One good aspect of Andrews' test is given in his claim that his test is less sensitive to outliers, and by implication to distributions with heavier tails than the normal. Andrews supports his claims by analysis of Box and Cox's two examples using his test for both of the Box and Cox examples and also for the biological example with one additional outlier added. His test is affected much less than the maximum likelihood test by addition of the outlier. One purpose of this work is to construct a more formal study of Andrews' conjecture by making an analysis using heavier-tailed distributions than the normal distributions from Box and Cox's examples.

In the paper by Atkinson, a comparison of three tests was given. The three tests were the Box and Cox and Andrews tests and another test derived by Atkinson. Atkinson decided to consider another test for two reasons: he wanted a test which was easy to compute and had higher power than the others, and also a test which did not neglect the remainder as the Andrews test did.

Atkinson expressed his test in the form which follows, using the transformed observations $\underline{z}^{(\lambda)}$ given in (5):

$$T_D = -\frac{\underline{z}^{(\lambda)}{}' A\underline{w}^{(z)}}{s_z \{\underline{w}^{(z)}{}' A\underline{w}^{(z)}\}^{\frac{1}{2}}} \tag{12}$$

where $\underline{w}^{(z)} = \partial\underline{z}^{(\lambda)}/\partial\lambda$ and s_z is an estimate of the variance σ_z^2 of the values $\underline{z}^{(\lambda)}$. The test he derived was a form of the locally most powerful test.

In order to compare the three test statistics T_L, T_A and T_D, Atkinson performed a series of simulations using the model given in Box and Cox's biological example. To determine the power of these tests, simulations were performed using different values of λ and the percentage of tests which were significant in each case was counted. He also gave a plot of these results, which indicated that Andrews' test T_A was much less powerful than the other two, especially at large distances from the true value of λ, but that the other two tests were similar in power.

It is questionable how good the results in Atkinson's paper are because he only considers one numerical example. He mentions this problem briefly, but since he has only done this one example, the conclusions must be based on the results of his example. Later in this paper, the results of further simulations which were performed in this manner using distributions other than normal will be given.

All of the tests which have thus far been considered were constructed on the assumption that for some λ, the transformed observations $\underline{y}^{(\lambda)}$ will follow a normal distribution. In a later paper by Hinkley (1975), robust analysis

was used to find another way to estimate λ. In this paper,
he did not assume any distribution for the transformed ob-
servations. He wished to find a value of λ for which the
transformed observations had a symmetric distribution.

If there are n independent and identically distributed
random variables $Y_1, \ldots, Y_n$, then the value of λ for which
the p and $1-p$ quantities are symmetric about the median is
the value that is desired. Since this will be expressed
in terms of the ordered values of $Y_1, \ldots, Y_n$, they will
be denoted by $X_1 \leq \ldots \leq X_n$. The value of λ that is desired
is the value for which

$$\tilde{X}^\lambda - X_r^\lambda = X_{n-r+1}^\lambda - \tilde{X}^\lambda \qquad (13)$$

where $r = [np]$ and $\tilde{X}$ is the median of the random variables.
The two solutions to this equation are $\lambda = 0$ and another
solution which Hinkley calls T. He excludes the value $\lambda = 0$
unless $\tilde{X}/X_r = X_{n-r+1}/\tilde{X}$ and he also rewrites the equation
as

$$(X_r/\tilde{X})^\lambda + (X_{n-r+1}/\tilde{X})^\lambda = 2. \qquad (14)$$

Hinkley states that the estimate T of λ has an asymptotic
normal distribution and he derives the asymptotic variance
of T.

In his discussion, Hinkley also states that problems
may arise when more complex models are used. He refers to
the Box and Cox biological example and states that different
estimates of λ may be found according to which sets of cell
means are examined. This is a large problem because most

models that we wish to analyze will be similar to the Box and Cox example. We need an estimate of λ which can also be used in these cases. Therefore the Hinkley estimate is only useful in certain simplified cases, and not in more complex cases.

We see that, of the four estimates and test statistics which we have considered, the evidence given to date indicates that the Atkinson test would probably be preferred over the other three. It is easier to compute than the maximum likelihood statistic, possibly more powerful than Andrews' test statistic and useful in more cases than the Hinkley estimate.

In this paper, the conjectures made by the other authors will be analyzed further. We have seen that problems may occur when distributions with heavier tails than the normal distribution are considered. The results of simulations performed using such distributions will be given and the power of the Andrews and Atkinson test statistics will be further considered.

CHAPTER III

INFLUENCE CURVES

The influence curve is a useful method of representing
how the behavior of a single observation affects an esti-
mator. It indicates how this single observation, which may
be an outlier, changes the value of the estimator, so it
is a measure of robustness. It is actually an expression
of the first derivative of an estimator, evaluated at a
certain distribution.

The influence curve will be denoted IC (x; T,F),
where T is the estimator in which we are interested and F
is the distribution at which it is evaluated. Let $\delta_x(y)$
be the function defined by $\delta_x(y) = \begin{cases} 0 \text{ for } y < x \\ 1 \text{ for } y \geq x \end{cases}$

If we view T as a functional depending on F, and denote it
T (F), the influence curve is defined in Hampel (1974) as

$$IC(x; T, F) = \lim_{\epsilon \downarrow 0} \{T(1-\epsilon)F + \epsilon\delta_x \ - T(F)\}/\epsilon \ . \tag{1}$$

Thus it is evident that the influence curve is the first
derivative of the estimator T at the distribution F.

If F_n is the empirical distribution function based
on a sample $X_1, \ldots, X_n$, the behavior of an estimator
$T(F_n)$ is described by

$$n^{\frac{1}{2}} \left| T(F_n) - T(F) - \sum_{i=1}^{n} IC(X_i; T, F) \right| \xrightarrow{p} 0, \tag{2}$$

and thus

$$n^{\frac{1}{2}}(T(F_n) - T(F)) \xrightarrow{L} N(0, \int(IC(x;T,F))^2 dF(x)). \tag{3}$$

Therefore it is evident from (2) that the influence curve describes the "influence" of a particular observation on $T(F_n)$.

A simple example of an influence curve, given in Hampel, is the influence curve of the arithmetic mean. $T = \int x dF(x)$ evaluated at any distribution F which has a finite first moment. If the mean of F is μ, then the influence curve is

$$IC(x; T, F) = \lim_{\varepsilon \downarrow 0} [(1-\varepsilon)\mu + \varepsilon x - \mu] / \varepsilon$$
$$= x - \mu.$$

Thus the influence of a point x on the arithmetic mean is a simple linear function of the point x, also depending on the mean μ of the distribution. This influence curve is unbounded, which implies that the arithmetic mean is not a robust estimator. A plot of this influence curve is given in Figure 1. Plots are also given of several other robust estimators in the following figures. These robust estimators will fall into two classes: trimmed means and M- estimators. Most of the results will be taken either from a book by Andrews and several others (1972) or a paper by Carroll and Wegman (1975).

One simple robust estimator of interest is the trimmed mean. The α - trimmed mean (for $0 < \alpha < \frac{1}{2}$) is found by ordering the observations in a sample, deleting the αn smallest and αn largest observations, and finding the arithmetic mean of the rest. The median of the sample is

seen to be the .50-trimmed mean. To find the influence
curve of the trimmed mean, we need the expression for the
α trimmed mean of any distribution F, which is

$$\int_\alpha^{1-\alpha} F^{-1}(t)\,dt /\quad (1-2\alpha) \qquad \text{(from Hampel)}.$$

The influence curve for the α- trimmed mean in the special
case of F being a symmetric distribution is given by

$$IC\ (x;T,F) = \begin{cases} F^{-1}(\alpha)/(1-2\alpha) & \text{for } x < F^{-1}(\alpha) \\ x/(1-2\alpha) & \text{for } F^{-1}(\alpha) \le x \ F^{-1}(1-\alpha) \\ F^{-1}(1-\alpha)/(1-2\alpha) & \text{for } x < F^{-1}(\alpha). \end{cases}$$

If the distribution F is asymmetric, the expression for the
influence curve is more complicated. For $\alpha = \frac{1}{2}$, the median,
assuming that F has a density f which is symmetric about
zero, the influence curve is

$$IC\ (x;\ T,\ F\ ) = \frac{\text{sign }(x)}{2f(0)} \qquad (5)$$

Thus, for the trimmed mean, the influence curve is bounded,
so the trimmed mean is a robust estimator. A plot of the
influence curve of the .10 - trimmed mean is given in Figure
2.

Another class of robust estimators of interest is the
class of M- estimators, which show very good robustness
properties. As given in Carroll and Wegman, M - estimators
are solutions, denoted by T, of an equation of the form

$$\sum_{j=1}^{n} \Psi\ \frac{(x_j - T)}{s} = 0\ , \qquad (6)$$

where ψ is an odd function and s is a scale estimate. The
estimate s can either be found independently or from an
equation of the form

$$\sum_{j=1}^{n} \frac{\chi(x_j - T)}{s} = 0 , \tag{7}$$

where χ is an even function. For Huber estimates, choose

$$\psi(x;k) = \begin{cases} -k & x < -k \\ x & -k < x < k \\ k & x > k \end{cases} , \tag{8}$$

use a specified function χ in (7), and solve simultaneously for T and S.

The same ψ function has also been used (by Hampel) along with $s = med|x_i - 50\%| /.6754$ to give a different estimator T found from solving equation (4). These estimators depend on the value of k which is chosen. The influence curve of the general M-estimator in the case where F is symmetric is given in Andrews et al, along with the statement that the influence curve is much more complicated in the asymmetric case. The influence curve is proportional to the function $\psi(x; k)$. A plot of the influence curve of the M-estimator with $k = 1.5$ is given in Figure 3.

The influence curves of the trimmed means and the Huber estimators, which one notes have the same general shape, both give some influence to large observations. Hampel proposed an estimator T which gave zero influence to large observations. He used the median of the absolute deviations from the median, which he called the median deviation, as his scale estimate and chose ψ to be

$$\psi(x; a, b, c) = \begin{cases} |x| & 0 \le |x| < a \\ a & a < |x| < b \\ \frac{c - |x|}{c - b} a & b \le |x| < c \\ 0 & |x| \ge c . \end{cases} \tag{9}$$

These estimates, which are called hampels, depend on the value of a, b and c which are chosen. Since $\overset{\psi}{\cdot}$ is zero for $|x|$ greater than a given constant, zero influence is given to an observation with $|x|$ greater than that constant. A plot of the influence curve of the hampel estimator with a = 2.5, b = 4.5 and c = 9.5 is given in Figure 4.

These estimators of location are useful for illustrating the idea of influence curves, because the results are somewhat simple. One important point to notice is that the expressions for the influence curves become much simpler when the underlying distribution F is symmetric. In the first part of the paper, power transforms which transformed data to normality were studied. It will now be useful to examine influence curves of the estimators of λ which were given there.

First we will consider the influence curve of the Box and Cox estimate of λ in the location problem. The Box and Cox estimate is the maximizing value of the log likelihood given in equation (2) of the previous chapter. In an unpublished work by Carroll, the influence curve of this estimate is derived. (This derivation is given in the Appendix.) The results are separated into two cases, $\lambda > 0$ and $\lambda < 0$. If $\lambda > 0$, the influence curve is of the order $y^{2\lambda} \log y$ as $y \to \infty$ and of the order $\log y$ as $y \to 0$. If $\lambda < 0$, the results are reversed, giving order $\log y$ as $y \to \infty$ and $y^{2\lambda} \times \log y$ as $y \to 0$. Looking at the specific case $\lambda = -1$, the

influence is of the order $\dfrac{\log y}{y^2}$ for observations near zero, so this estimate should be sensitive to quite small observations if $\lambda = -1$. A plot of the influence curve of the Box and Cox estimate with $\lambda \cong -1$ is given in Figure 5.

Next, the Andrews and Atkinson estimates are considered. The influence curve does not exist in general for the Andrews estimate, but some information can be found for $\lambda = -1$. In this case, more influence is given to observations near zero than is given by the Box and Cox estimate. The same type of calculations are used to find the general influence curve for the Atkinson estimate. If we again look at the case $\lambda = -1$, the influence of an observation near zero is found to be of the order $\dfrac{(\log y)^2}{y^2}$; thus it is also more sensitive to small observations than the Box and Cox maximum likelihood estimate.

The Hinkley estimate is also considered. The results are not quite as complicated, so a general expression for the influence curve can be found. The influence curve is seen to depend on the derivative of the underlying distribution and the value which is chosen. It is a bounded function with three discontinuities, so it is not as sensitive to large observations (if $\lambda > 0$) or small observations (if $\lambda < 0$) as the other estimates. It is still not desirable over the other estimates, though, because the results it gives are not particularly realistic for more complicated models.
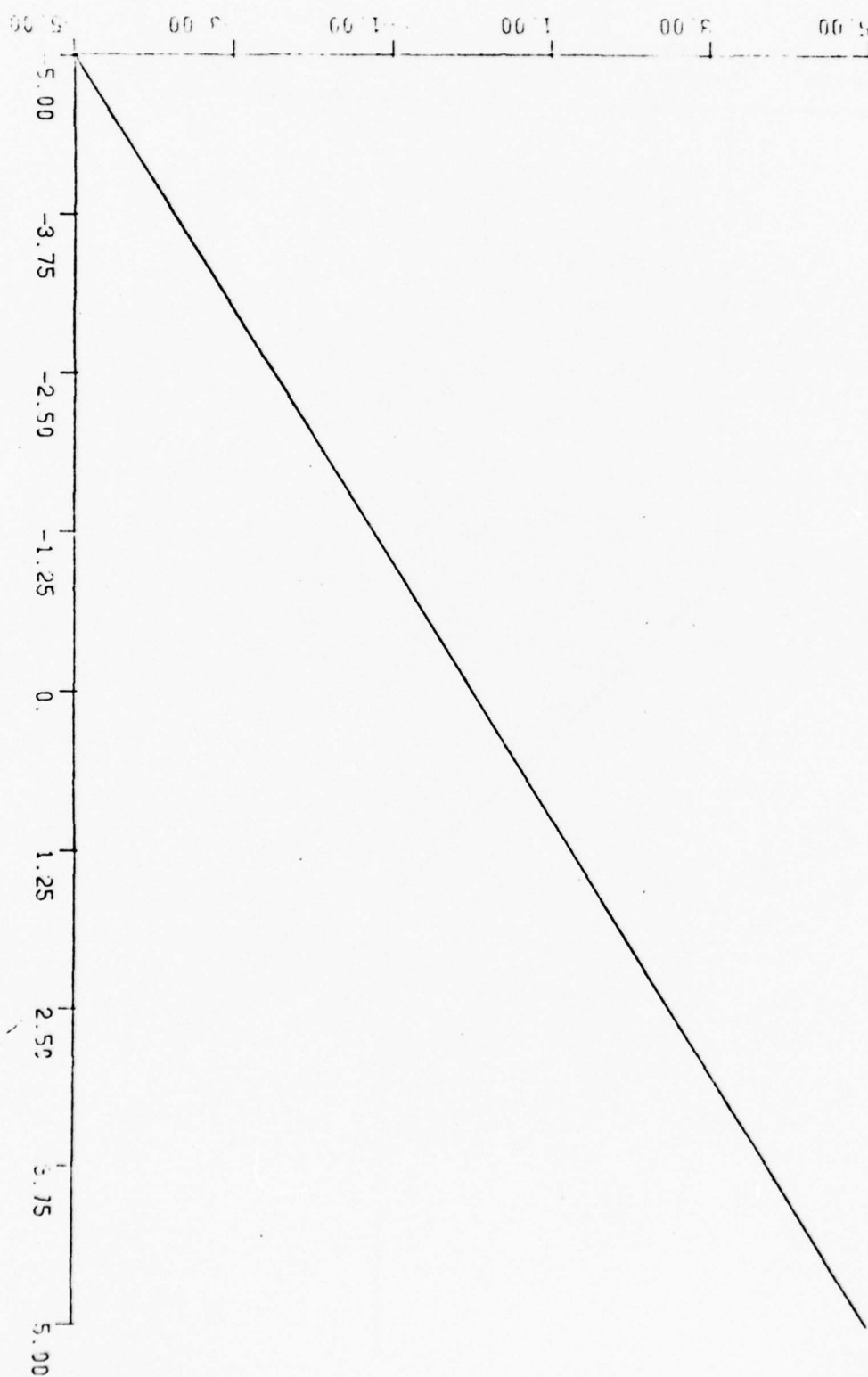
FIGURE 1. Influence curve for mean.

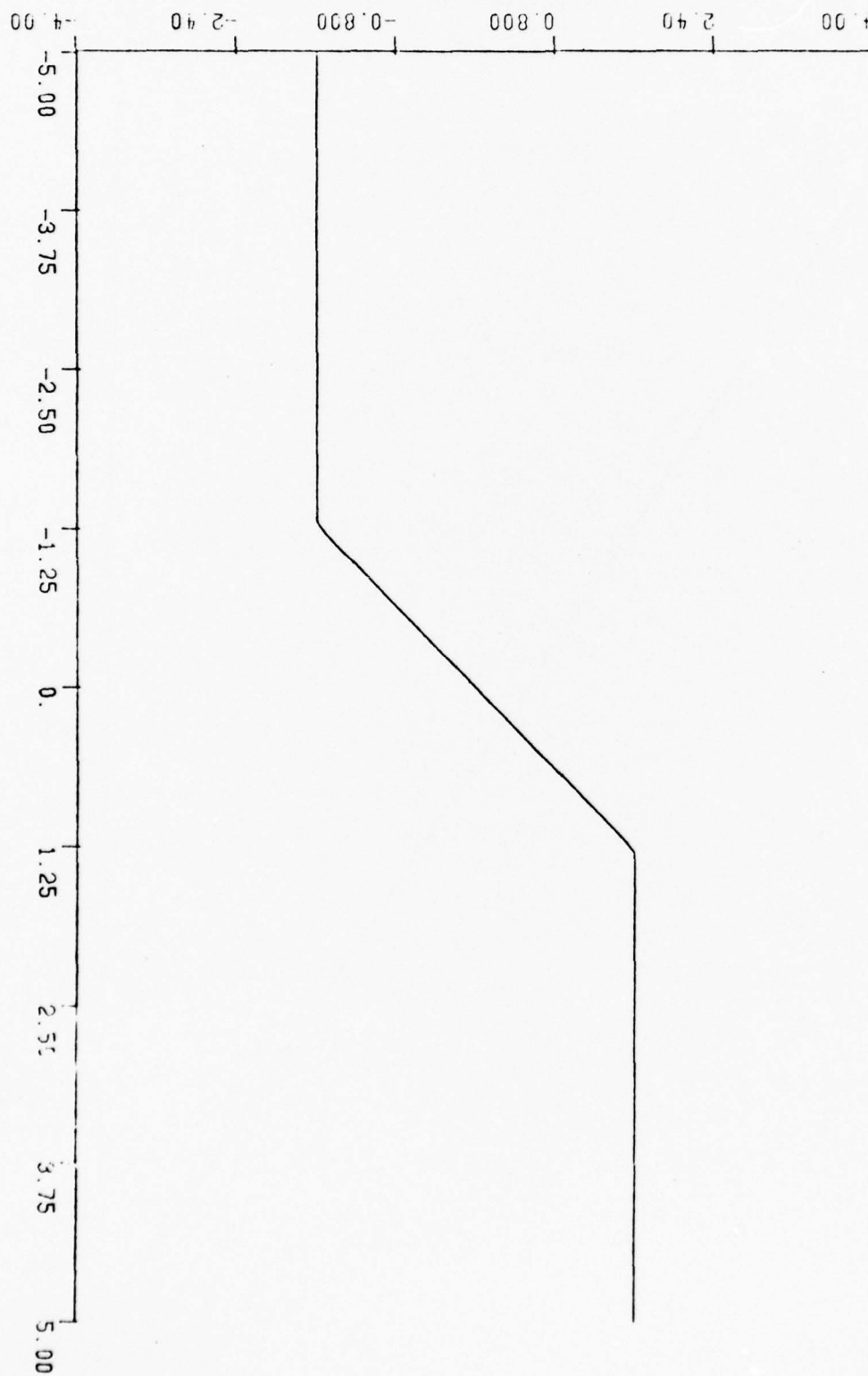FIGURE 2. Influence curve for 10% symmetrically trimmed mean.
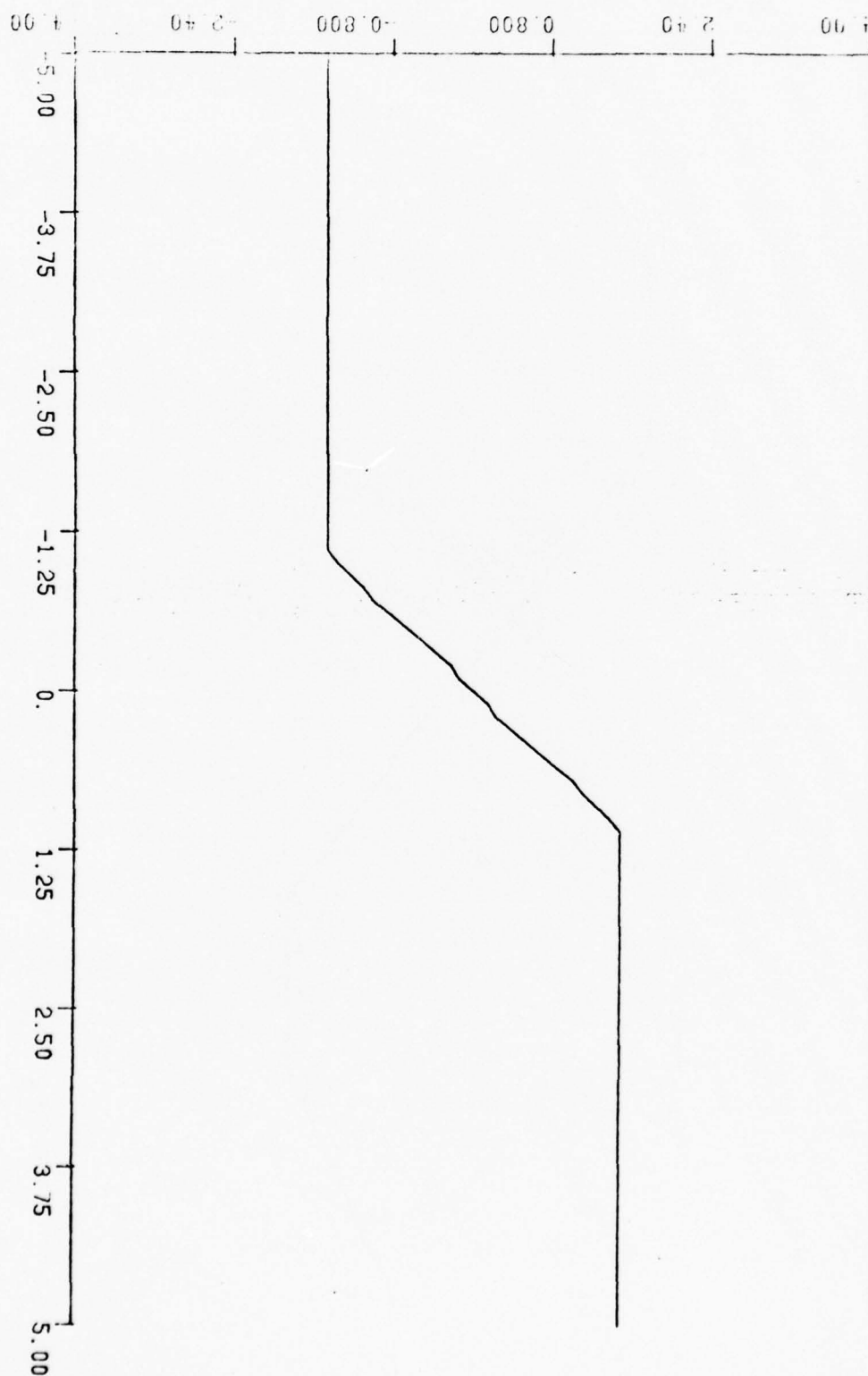
FIGURE 3. Influence curve for M estimate, k=1.5.

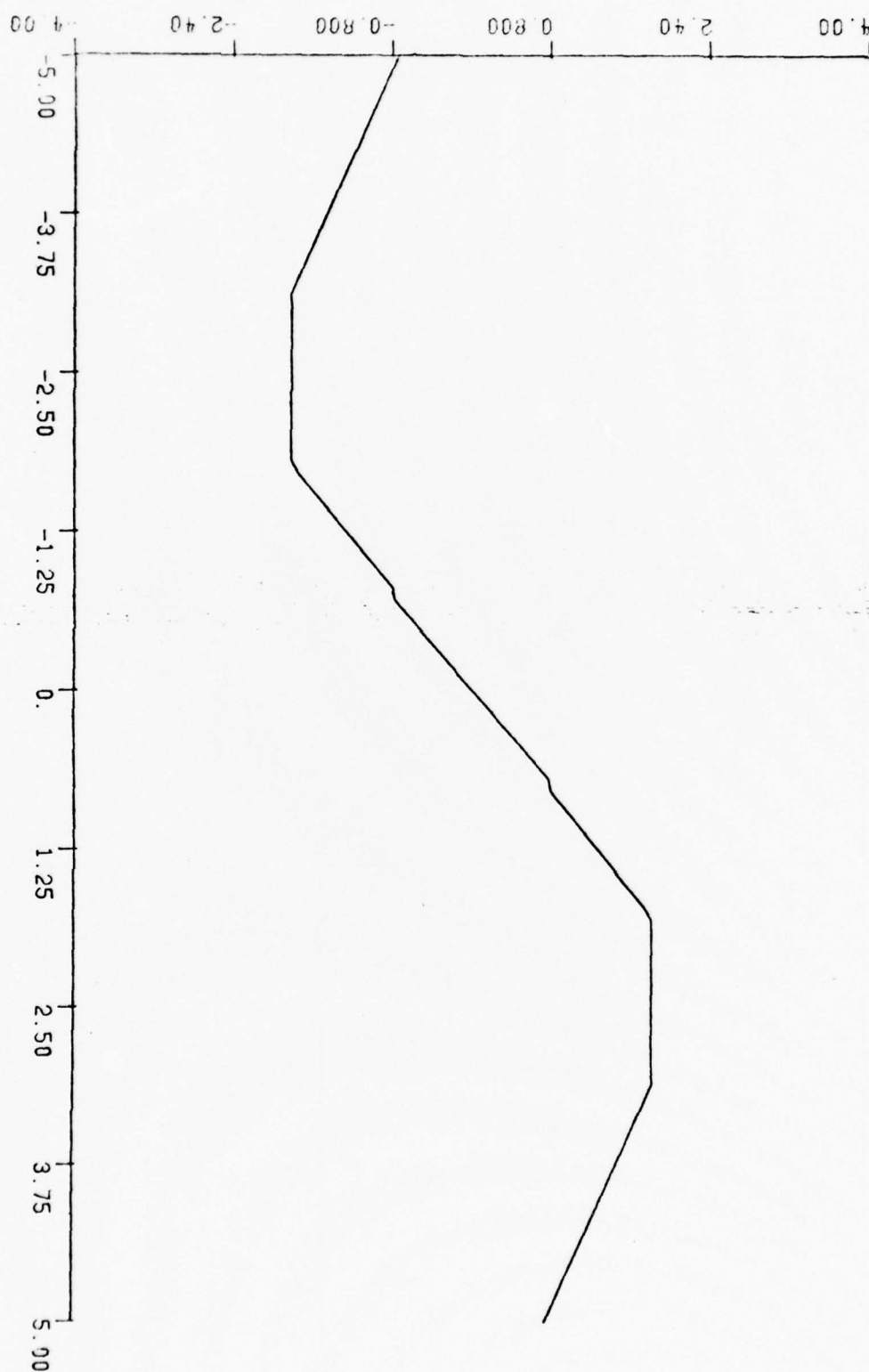FIGURE 4. Influence curve for hampel estimate,
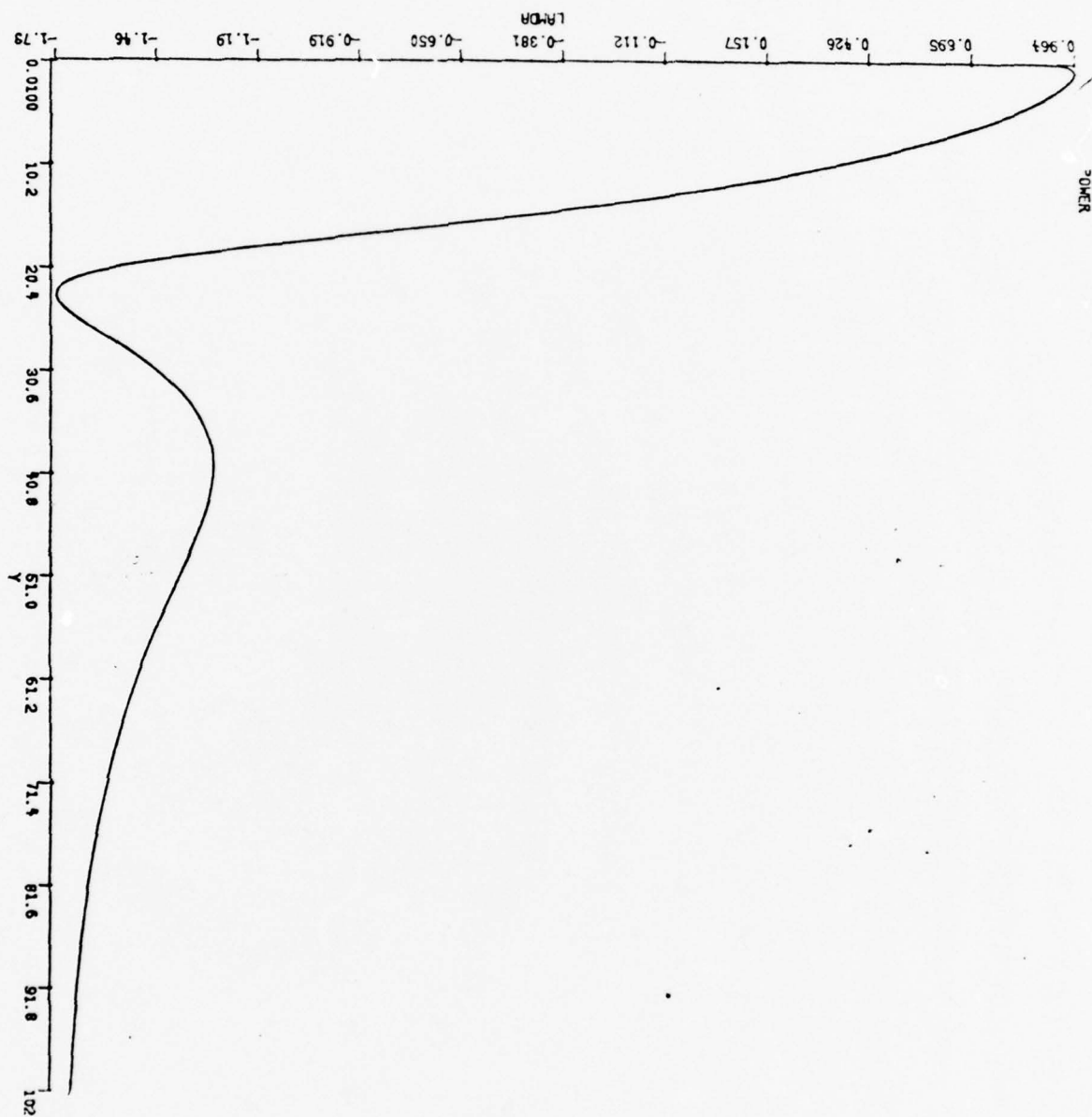a=2.5, b=4.5, c=9.5.

FIGURE 5. Influence curve for the Box and Cox estimate
with $\lambda \overset{\sim}{=} -1$.

CHAPTER IV

RESULTS OF SIMULATIONS

As stated previously, one of the main purposes of this paper
is to further study the power of two of the tests presented in Chapter
II by performing a series of simulations. These simulations were per-
formed using the same method as in the paper by Atkinson.

In Atkinson's paper, he describes simulations which were per-
formed to study the power of three of the tests which were described
in Chapter II. The three tests which he used were the Box and Cox
maximum likelihood test T_L, Andrews' exact test T_A and his own test
T_D. The simulations were based on the data from the biological example
in the Box and Cox paper. Here we chose to perform simulations using
only the Andrews test T_A and the Atkinson test T_D, because this made
the computations easier.

Atkinson's simulations were performed using the data from the
Box and Cox example to generate normally distributed data. In order
to study the robustness of the tests, we chose to generate random
variates from three different distributions. The first of these dis-
tributions was the normal, which was chosen to reproduce the results
given in Atkinson's paper. The other two distributions chosen have
thicker tails than the normal. The first of these was called the con-
taminated normal; this was 90 per cent $N(0,1)$ and 10 per cent $N(0,9)$.
The other was the double exponential, which as generated had variance

2. All three of these types of random variates were multiplied by a factor of $.5/\sqrt{2}$. The factor $1/\sqrt{2}$ was chosen to give the double exponential a variance of 1 and the factor .5 in the numerator was chosen to more closely imitate the work done by Atkinson. The resulting variances were .125 for the normal, .225 for the contaminated normal and .25 for the double exponential.

The means for all three of the different types of random variates were generated in the same way, directly from the Box and Cox biological data. The first step was to arrange the data into a 48x1 vector $\underline{y}$; then, the data were transformed into a vector $\underline{y}^{(\lambda)}$. Since we were interested in testing the hypothesis $\lambda=-1$, this value was chosen in transforming the observations. The means used were the predicted means $\hat{\underline{y}}^{(\lambda)}=X(X'X)^{-1}X'\underline{y}^{(\lambda)}$.

In order to study the power of the two tests T_A and T_D, the simulations were repeated with different values of λ. To accomplish this, the cell means were generated in the method above with $\lambda=-1$, transformed back to the original scale by taking the inverse of the Box and Cox transformation $\underline{y}=1/(1-\underline{y}^{(\lambda)})$ and then transformed again using the new value of λ. The values of λ used were $\lambda=-1.5,-1,-.5,-.05$ and .4. The results of these 15 simulations are given on the following pages in Tables 1,2 and 3, expressed as the number of 200 simulations which resulted in significant test statistics for the tests T_A and T_D for all three distributions. Three different plots are also given in Figures 6, 7 and 8, one for each different distribution. As stated in the paper by Atkinson, the slope of these plots indicates the power of the tests.

The results of these simulations agree with Atkinson's results for the normal case, since the power of the test T_D is greater than the power of the test T_A for all 5 values of λ. The results for the

other two distributions are partly consistent with the normal case,
because the power of T_D is larger than the power of T_A at all values
of λ for both different distributions. Away from the null hypothesis,
though, there is a loss of efficiency since the power is lower for
both T_D and T_A than in the normal case. There is also a loss of validity
at the null hypothesis, because the intended 5 per cent tests become
closer to 30 per cent for the Atkinson test and 10 per cent for the
Andrews test.

TABLE 1

Power of the two tests for testing $\lambda = -1$. Data generated from the normal distribution with $\lambda = k$. Number out of 200 simulations significant at the 5 per cent level.

k	T_A +	T_A -	T_D +	T_D -
-1.5	0	103	0	181
-1	5	2	9	8
- .5	91	0	175	0
- .05	184	0	200	0
.4	185	0	200	0

TABLE 2

Power of the two tests for testing $\lambda = -1$. Data generated from the contaminated normal distribution with $\lambda = k$. Number out of 200 simulations significant at the 5 per cent level.

k	T_A +	T_A -	T_D +	T_D -
-1.5	0	64	0	185
-1	7	11	7	56
- .5	46	0	125	34
- .05	111	0	165	34
.4	116	0	173	22

TABLE 3

Power of the two tests for testing $\lambda = -1$. Data generated from the double exponential distribution with $\lambda = k$. Number out of 200 simulations significant at the 5 per cent level.

	T_A			T_D	
k	+	-		+	-
-1.5	0	59		0	180
-1	4	6		5	54
-.5	40	1		122	36
-.05	92	0		163	36
.4	105	0		166	27

FIGURE 6. Power of the two statistics for testing $\lambda=-1$ using normally distributed data. Proportion of 200 simulations significant at the 5 per cent level. A denotes T_A- Andrews' exact test and D denotes T_D- Atkinson's test.
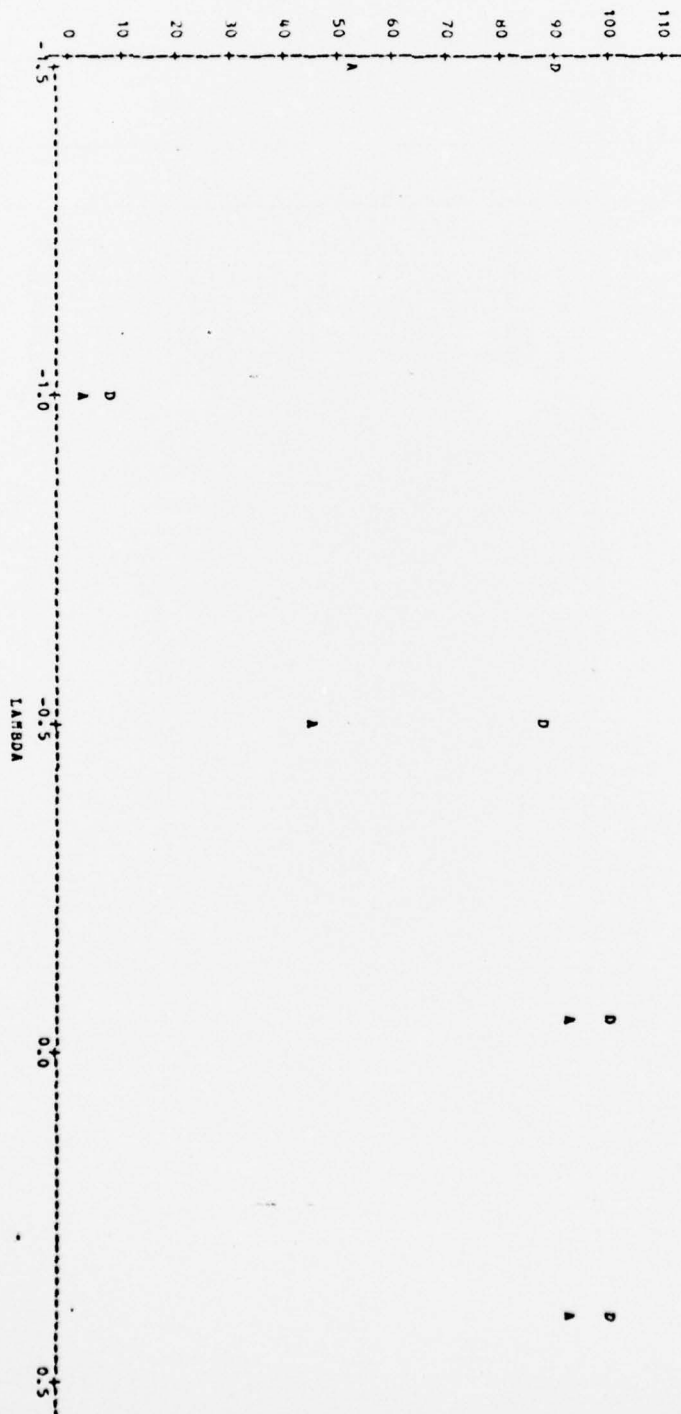
FIGURE 7. Power of the two test statistics for testing λ=-1
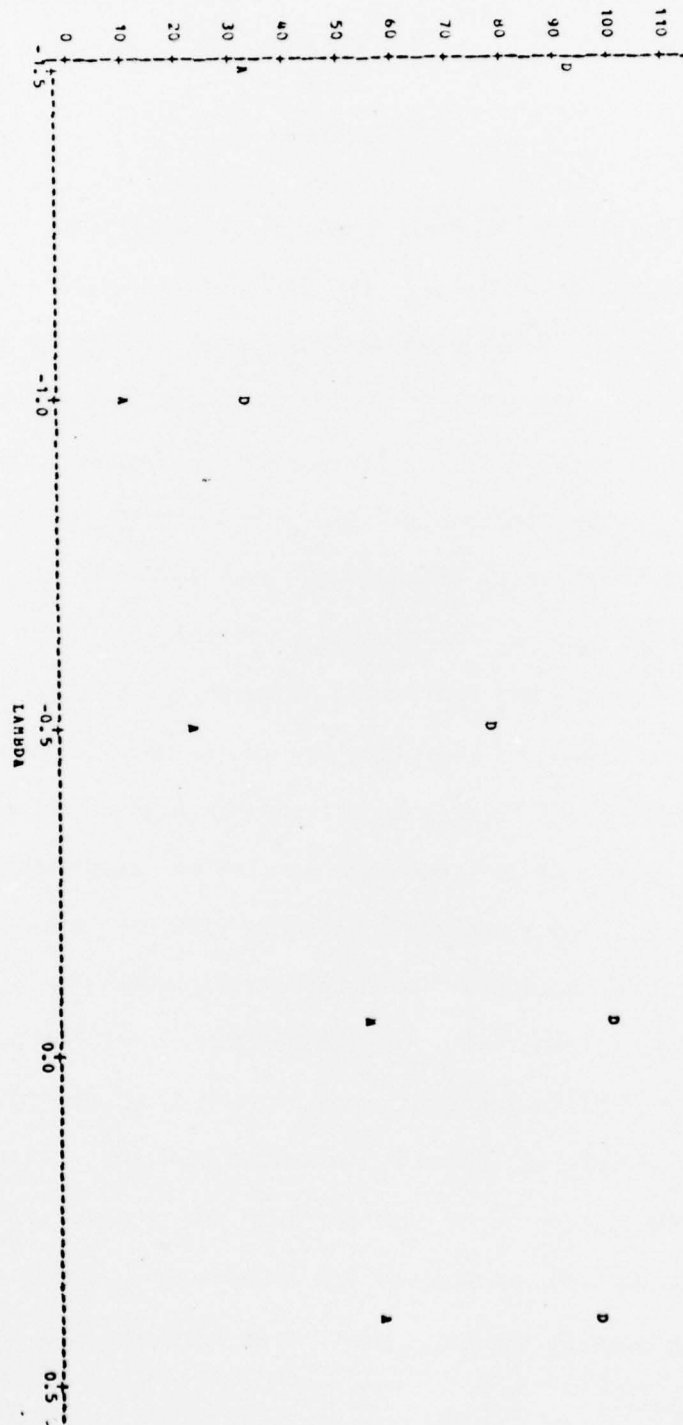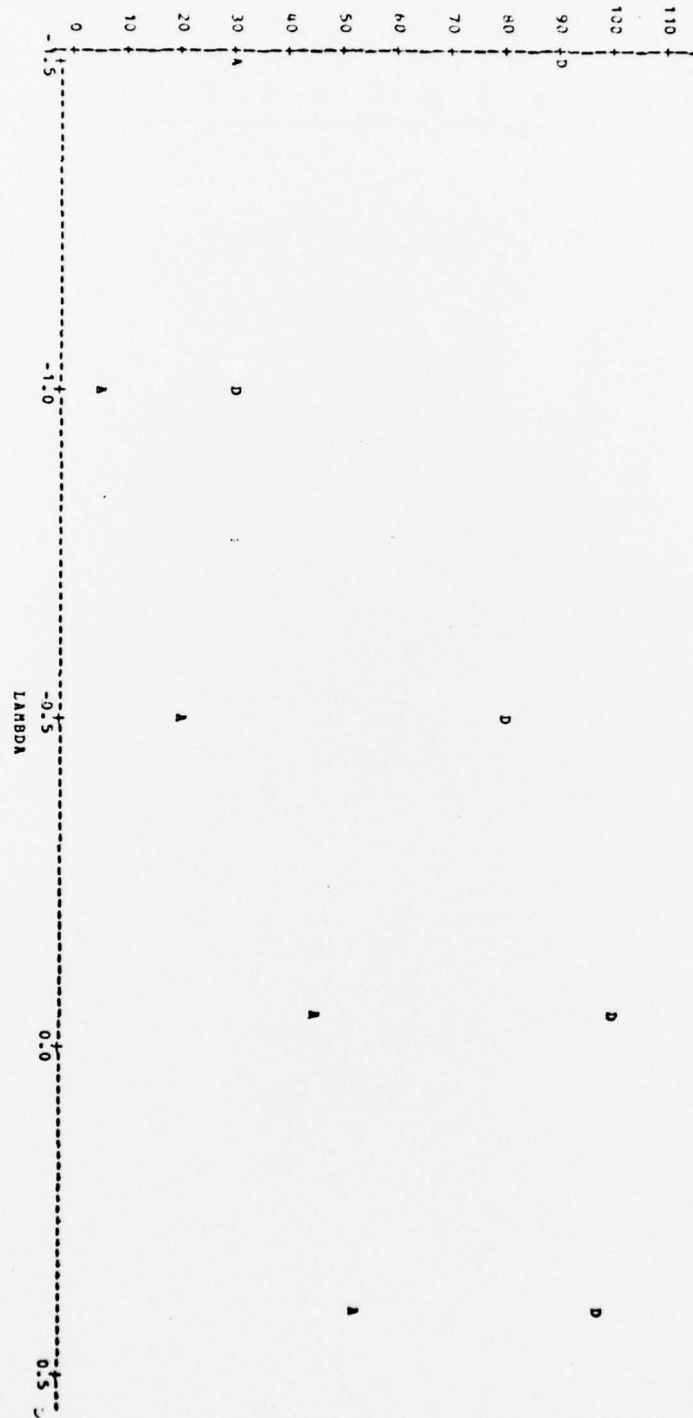using contaminated normal data.

FIGURE 8. Power of the two test statistics for testing λ=-1
using double exponential data.

CHAPTER V

CONCLUSION

In the first part of the paper, four different estimates for the
optimal value of λ were studied: the Box and Cox, Andrews, Atkinson
and Hinkley estimates. Test statistics for testing $\lambda=\lambda_0$ were also
derived for the Box and Cox, Andrews and Atkinson cases. The results
of a numerical example in the paper by Atkinson gave an indication of
the power of the three tests T_L-the Box and Cox test, T_A-the Andrews
test and T_D-the Atkinson test for normally distributed data. Atkinson
concluded that his statistic T_D was similar in power to the statistic
T_L and that both were greater in power than the statistic T_A.

A series of simulations was performed to expand on the results
given in the paper by Atkinson. The purpose of these simulations was to
study the robustness, as well as the power, of the tests T_A and T_D.
Whereas Atkinson used only normally distributed data, the simulations
here included normally distributed data, data from a contaminated normal
distribution and double exponential data. The results here indicated
that the power of T_D is greater than the power of T_A for all three types
of distributions. For the contaminated normal and double exponential
distributions, though, the Atkinson test shows an extreme loss of
validity at the null hypothesis and the Andrews test shows a slight loss
of validity. Away from the null hypothesis, the Andrews test shows an

extreme loss of efficiency and the Atkinson test shows a slight loss of efficiency.

Since the contaminated normal and double exponential distributions both have heavier tails than the normal, some problems were expected when these distributions were considered. Since the Andrews test is an exact t-test, the loss of efficiency would be expected, because the usual t-tests display this loss of efficiency away from the null hypothesis. The loss of validity of the Atkinson test should also be expected from examining the influence curve, because quite large influence is given to both large and small observations when $\lambda=-1$ (which is the null hypothesis). Thus the conclusion is that the two tests are not very robust to heavier tailed distributions than the normal, because of the above mentioned losses of validity and efficiency.

APPENDIX

The Influence Curve for the Box and Cox Estimate

The Box and Cox estimate is the value which maximizes the log
likelihood function

$$L \max(\lambda) = (\lambda-1) \frac{1}{n} \Sigma \log y_i$$
$$- \frac{1}{2} \log\{\frac{1}{n} \Sigma y_i^{(\lambda)^2} - (\frac{1}{n} \Sigma y_i^{(\lambda)})^2\} \, .$$

To maximize this function, we take derivatives with respect to λ and
let $n \to \infty$. Evaluating this derivative at the "true" value λ_0, we obtain

$$0 = \lim_{n\to\infty} \frac{\partial}{\partial\lambda} L \max(\lambda)\Big|_{\lambda=\lambda_0}$$

$$= \lim_{n\to\infty} (-n \underline{y}^{(\lambda)'} A \underline{u}^{(\lambda)} + \frac{n}{\lambda} \underline{y}^{(\lambda)'} A \underline{y}^{(\lambda)}$$
$$+ (\underline{y}^{(\lambda)'} A \underline{y}^{(\lambda)}) \Sigma \log y_i)\Big|_{\lambda=\lambda_0}$$

$$= E \log \underline{y} - \frac{1}{S^2(F,\lambda_0)} \{E \underline{y}^{(\lambda_0)} \frac{\partial}{\partial\lambda} \underline{y}^{(\lambda_0)} - T(F,\lambda_0) \frac{\partial}{\partial\lambda} \underline{y}^{(\lambda_0)}\}$$

where

$$E (\frac{\underline{y}^{(\lambda_0)} - T(F)}{S(F)}) = 0$$

and

$$E[(\frac{\underline{y}^{(\lambda_0)} - T(F)}{S(F)})^2 - 1] = 0 \, .$$

Now, to compute the influence curve, we let $\lambda_0 = \lambda(F)$ for given F and
define the following equations:

$$\psi_1(y;T(F),S(F),\lambda(F)) = \frac{y^{(\lambda(F))} - T(F)}{S(F)}$$

$$\psi_2(y;T(F),S(F),\lambda(F)) = \frac{(y^{(\lambda(F))} - T(F))^2}{S^2(F)} - 1$$

$$\psi_3(y;T(F),S(F),\lambda(F)) = \log y - \frac{1}{S^2(F)}(y^{(\lambda(F))} - T(F)) \frac{\partial}{\partial\lambda} y^{(\lambda(F))} \ .$$

Then the functionals $T(F)$, $S(F)$ and $\lambda(F)$ are the solutions to the system of equations

$$\int \psi_i(y;T(F),S(F),\lambda(F))dF(y) = 0$$

for $i = 1,2,3$.

To derive the influence curve, we first need to define the distribution functions

$$F_\epsilon(y) = (1-\epsilon)F_0(y) + \epsilon\delta_x(y)$$

where

$$\delta_x(y) = \begin{cases} 0 & y < x \\ 1 & y \geq x \end{cases} \ .$$

Then the above system of equations becomes

$$\int \psi_i(y;T(F_\epsilon),S(F_\epsilon),\lambda(F_\epsilon))dF_\epsilon(y) = 0$$

which implies that

$$(1-\epsilon)\int \psi_i(y;T(F_\epsilon),S(F_\epsilon),\lambda(F_\epsilon))dF_0(y)$$

$$+ \ \epsilon\int \psi_i(y;T(F_\epsilon),S(F_\epsilon),\lambda(F_\epsilon))d\delta_x(y) = 0$$

or

$$(1-\epsilon)\int \psi_i(y;T(F_\epsilon),S(F_\epsilon);\lambda(F_\epsilon))dF_0(y)$$

$$+ \ \epsilon\psi_i(x;T(F_\epsilon),S(F_\epsilon),\lambda(F_\epsilon)) = 0 \ .$$

Next we take derivatives with respect to ϵ and evaluate at $\epsilon = 0$ to give

$$- \ \psi_i(x;T(F_0),S(F_0),\lambda(F_0))$$

$$= a_{1i}[\ \frac{\partial}{\partial\epsilon} T(F_\epsilon)\Big|_{\epsilon=0}] + a_{2i}[\ \frac{\partial}{\partial\epsilon} S(F_\epsilon)\Big|_{\epsilon=0}] + a_{3i}[\ \frac{\partial}{\partial\epsilon} \lambda(F_\epsilon)\Big|_{\epsilon=0}]$$

where the a_{1i}, a_{2i} and a_{3i} are coefficients which will be found from the above equations.

$$a_{11} = E - \frac{1}{S(F)} = -\frac{1}{S(F)}$$

$$a_{21} = E \frac{y^{(\lambda(F))} - T(F)}{-S^2(F)} = -\frac{1}{S(F)} E \frac{y^{(\lambda(F))} - T(F)}{S(F)} = 0$$

$$a_{31} = E \left. \frac{\partial}{\partial \lambda} y^{(\lambda)} \right|_{\lambda(F)} = E \left. \frac{y^\lambda \log y - y^{(\lambda)}}{\lambda} \right|_{\lambda(F)}$$

$$= E \frac{(y^{\lambda(F)} \log y) - T(F)}{\lambda(F)}$$

$$a_{12} = E \frac{-2(y^{(\lambda(F))} - T(F))}{S(F)} = 0$$

$$a_{22} = E -2 \left(\frac{y^{(\lambda(F))} - T(F)}{S(F)} \right)^2 \frac{1}{S(F)} = -\frac{2}{S(F)}$$

$$a_{32} = E \frac{2(y^{(\lambda(F))} - T(F))}{S^2(F)} \left. \frac{\partial}{\partial \lambda} y^{(\lambda)} \right|_{\lambda(F)}$$

$$= \frac{2}{S^2(F)} E(y^{(\lambda(F))} - T(F)) \left(\frac{y^{\lambda(F)} \log y - y^{(\lambda(F))}}{\lambda(F)} \right)$$

$$a_{13} = E \frac{(\partial/\partial \lambda) y^{(\lambda(F))}}{S^2(F)} = \frac{(\partial/\partial \lambda) E y^{(\lambda(F))}}{S^2(F)} = 0$$

$$a_{23} = \frac{2}{S^3(F)} E(y^{(\lambda F)} - T(F)) \left(\frac{\partial}{\partial \lambda} y^{(\lambda(F))} \right)$$

$$a_{33} = \frac{1}{S^2(F)} \left[T(F) \frac{\partial}{\partial \lambda} y^{(\lambda(F))} - y^{(\lambda(F))} \frac{\partial}{\partial \lambda} \left(\frac{y^{\lambda(F)} \log y}{\lambda(F)} - \frac{(y^{\lambda(F)} - 1)}{\lambda^2(F)} \right) \right.$$

$$\left. - y^{(\lambda(F))} \frac{\partial}{\partial \lambda} \left(\frac{y^{(\lambda(F))} \log y}{\lambda^0(F)} - \frac{(y^{\lambda(F)} - 1)}{\lambda^2(F)} \right) \frac{\partial}{\partial \lambda} y^{(\lambda(F))} \right] .$$

Thus the above equations become

$$\begin{bmatrix} -\dfrac{1}{S(F)} & 0 & a_{31} \\ 0 & -\dfrac{2}{S(F)} & a_{32} \\ 0 & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} \dfrac{\partial}{\partial\epsilon} T(F_\epsilon) \\ \dfrac{\partial}{\partial\epsilon} S(F_\epsilon) \\ \dfrac{\partial}{\partial\epsilon} \lambda(F_\epsilon) \end{bmatrix} = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} \quad .$$

To get the solutions of the matrix equations and find the influence curve, take the inverse of the matrix to give

$$\begin{bmatrix} \dfrac{\partial}{\partial\epsilon} T(F_\epsilon) \\ \dfrac{\partial}{\partial\epsilon} S(F_\epsilon) \\ \dfrac{\partial}{\partial\epsilon} \lambda(F_\epsilon) \end{bmatrix} = - \begin{bmatrix} -\dfrac{1}{S(F)} & 0 & a_{31} \\ 0 & -\dfrac{2}{S(F)} & a_{32} \\ 0 & a_{23} & a_{33} \end{bmatrix}^{-1} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} \quad .$$

Now we see that the influence curves are functions of the ψ_i.

$$\psi_1(y;T(F_0),S(F_0),\lambda(F_0)) = \frac{y^{(\lambda(F))} - T(F)}{S(F)}$$

is of order y^λ,

$$\psi_2(y;T(F_0),S(F_0),\lambda(F_0)) = \frac{(y^{(\lambda(F))} - T(F))^2}{S^2(F)} - 1$$

is of order $y^{2\lambda}$ and

$$\psi_3(y;T(F_0),S(F_0),\lambda(F_0)) = \log y - \frac{1}{S^2(F)}(y^{(\lambda(F))} - T(F))\frac{\partial}{\partial\lambda} y^{(\lambda(F))}$$

is of order $y^{2\lambda}\log y$.

Therefore, if $\lambda > 0$, the influence curve is of the order $y^{2\lambda}\log y$ as $y \to \infty$ and order $|\log y|$ as $y \to 0$. If $\lambda < 0$, the results are reversed to give order $\log y$ as $y \to \infty$ and order $|y^{2\lambda}\log y|$ as $y \to 0$. An indication of these results is given in the plot of the influence curve for $\lambda \cong -1$ which is given in Figure 5.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
AFOSR-TR-77-0712		

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
A STUDY OF POWER TRANSFORMS	Interim
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Robert A. Fondren	AFOSR-75-2796

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
University of North Carolina Department of Statistics Chapel Hill, NC 27514	61102F 2304/A5

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332	May 1977
	13. NUMBER OF PAGES
	37

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

normal distributions, power transformation, influence curves, simulation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In most practical situations, the data one has do not follow the normal distribution; if the data were normal, one could use all of the procedures of multivariate analysis which apply to normally distributed data. In this paper, a routine method for making the data follow the normal distribution more closely is discussed; this method is to transform the data.

DD FORM 1473 1 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE   UNCLASSIFIED

BIBLIOGRAPHY

Andrews, D. F. (1971). A note on the selection of data trans-
        formations. Biometrika (58), 249-54.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J.,
        Rogers, W. H., and Tukey, J. W. (1972). Robust Esti-
        mates of Location: Survey and Advances. Princeton
        University Press, Princeton, N.J.

Atkinson, A. C. (1973). Testing Transformations to Normality.
        J. R. Statist. Soc. B (35), 473-9.

Box, G. E. P. and Cox, D. R. (1964). An Analysis of Trans-
        formations. J. R. Statist. Soc. B (26), 211-52.

Carroll, R. J. (1977). Influence Curves for Power Transforms.
        Unpublished manuscript.

Carroll, R. J. and Wegman, E. J. (1975). A Monte Carlo Study
        of Robust Estimators of Location. Institute of
        Statistics Mimeo Series #1040. University of North
        Carolina at Chapel Hill.

Draper, N. R. and Cox, D. R. (1969). On Distributions and
        Their Transformations to Normality. J. R. Statist.
        Soc. B (31), 472-6.

Hampel, F. R. (1974). The Influence Curve and Its Role in
        Robust Estimation. J. Am. Statist. Assoc. (69),
        383-93.

Hinkley, D. V. (1975). On Power Transformations to Symmetry.
        Biometrika (62), 101-11.

Milliken, G. A. and Graybill, F. A. (1970). Extensions of
        the General Linear Hypothesis. J. Am. Statist. Assoc.
        (65), 797-807.

Schlesselman, J. (1971). Power Families: A Note on the Box
        and Cox Transformation. J. R. Statist. Soc. B (33),
        307-11.